Scientific Summary

Bartha Maria Knoppers
Project Director
Centre de recherche en Droit Public
Université de Montréal
3101, Chemin de la tour
Montréal, (QC) H3T 1J7
knoppers@droit.umontreal.ca
Phone: +1 (514) 343-6714
Fax: +1 (514) 343-2122

Summary of Workshops

Focus areas for discussions at the workshops were:

i) Similarities and differences among the projects
ii) Key areas of collaboration
iii) Emerging issues, next steps, further harmonization

Working Group 1: Socio-demographic/health questionnaires

Leader: Professor Andres Mestpalu

Content of Questionnaires

- Sharing of questionnaires and protocols would be helpful for harmonization
- Harmonization should occur when possible, on actual questions or on domains, to facilitate validation of results

Scope of questions

- Designing one-stop surveys that allow pooling of subgroups that would need to be asked more specific questions in the future
- Pooling of data from different subgroups could enhance statistical power

One stop or multiple surveys

- One stop model is more common and cost efficient
- Multiple surveys measure varying exposures and increase opportunities for harmonization by allowing administrators to go back to participants and ask more specific questions

Administration

- Self-completion of questionnaires reduces cost and decreases interviewer variability
- Self completed questionnaires allow for better answers to non-clinical questions, while interviews by a GP allow for better recording of health conditions
- If self administration- by internet, e-mail, paper, phone helpline?

Pooling/meta-analysis

- Comparing different populations
- Maximizing statistical power
- Issue of whether to pool data at the individual level, or undertake a meta-analysis of the aggregated data from individual databases

Use of macro-environmental indicators

- Collection of scientifically valid and harmonized information on environmental factors is inexpensive and important for determining relationship between environment and disease
- Could use weather, pollution, water etc. data

Questionnaires must be validated

- Each question should be based on existing literature or guidelines so that the purpose of each question asked can be revisited

- Need to monitor quality assurance of questionnaires

Harmonization of instruments

- Instruments need to be validated for reliability and sensitivity at the local level, and the results presented to the consortium
- Use of already validated instrumentation

Other issues

- Population issues to consider including variability in age, random vs. population volunteers
- Data must be truly representative of all spheres of the population
- Biomarkers and measures would be better than questions for biological variables such as diet, physical measures etc., underlining need for standardizing the types of tests to be conducted
- Quality assurance standards needed
- Team should remain aware of groups not in the consortium that have population banks
- Strengths of using multiple instruments (e.g. questionnaires and biomarkers) to obtain maximum information

Estonian Project update:

- Team will take feedback from pilot project (over 2000 patients) and modify questionnaires accordingly
- GPs administered questionnaires
- Preliminary results suggest that participants were willing to participate, but felt bothered when asked for consent for each subsequent use

Working Group 2: Physical/physiological/biological/biochemical measures

Leader: Professor Leena Peltonen

Differences

- Study design between projects
- Recruitment methods (GP vs. random vs. twins), participation rates and the possibility to follow-up
- How much data should be recorded and which parameters and phenotypes should be measured, taking into account the tradeoff between amount of data collected and response rate
- Although there were many differences among the project, this did not imply that information could not be shared, nor did it impede exchange of experiences and difficulties during implementation of the project

Project	Stage of study	Length of study	Recruitment method	Samples	Number of Participants	Age range	Possibility of follow-up
CARTaGENE	Project development	At least 4 yrs	Randomized call	Double-coded	About 60 000	25-74	Yes*
Estonian Genome	Pilot project has begun	Ongoing Since 2001	Participants contacted by GP	Coded	Aiming for 1 million	?	Yes
GenomEUtwin	Project well underway	Ongoing	Twin cohort	Coded?	About 800 000 twins	?	Yes
UK Biobank	Project implementation	Samples monitored for 10 yrs	Randomly approached by GP	Coded	About 500 000	45-69	Yes

* Being considered: possibility of follow-up through governmental agencies and updating questionnaires. No access to individual medical records.

Similarities

- All population-based
- Composed of a large sample size
- Geographical anchors (important to determine)
- Public health perspective
- Clear ethical and social underpinnings
- Promote public endeavors
- Shared interest in sharing expertise and harmonizing elements such as SOPs, calculations and methodologies in sample collection, processing and storage

Key areas of collaboration

  - Harmonizing basic infrastructure
  - Measuring common and important parameters
  - Sharing expertise and tools
  - Sustaining ongoing communication
- Harmonizing data parameters relative to phenotypes to facilitate comparison of measures across data sets
- Core phenotypes in 'Federated Database' of P3G – importance of defining limitations may require one list of phenotypes common for P3G, and one list that may be more detailed
- Standardize structure of data that will be collected
- Main focus on parameters that involve public health
- Another important parameter is the environment
- Need to set strict guidelines for sample collection and storage (blood and DNA)

Create Working Groups for various aspects of collaboration

- WG 1: Comparison table on study design etc.
- WG 2: Phenotype questionnaires (health related) with criteria (why it is in, why it is out)
- WG 3: Physical traits
- WG 4: SOPs with biological samples (cQC)
- WG 5: Biochemical measures

Determining important phenotypes

- Major cause of morbidity/mortality
- Impact on public health
- Feasibility of measures
- Existing data- awareness of biases of existing data towards gender, screening results

Preliminary list of phenotypes:

  - Aging (cardiovascular disease, bone disease, cancer)
  - Arthritic disease
  - Autoimmune disease
  - Body composition (obesity)
  - Cancer
  - Diabetes
  - Eye disease

- Gastrointestinal disease
- Incontinence
- Mental illness/decline
- Metabolic syndrome
- Nutritional status
- Respiratory disorders, skin diseases, allergies
- Sensory evaluation
- Social and economic factors
- Toxicology
  - For some aspects, e.g. alcoholism, testing is more reliable than a questionnaire
  - Some phenotypes are age- or sex-specific, therefore measuring estrogen levels and age of menopause could also be useful
  - Special phenotypes to consider (may require special equipment):
    - Multifrequency bioimpedance (lean mass, fat mass, body composition) to measure ageing of the body (5 minutes)
    - Bone heel ultrasound, useful for studying ageing and osteoporosis (5 minutes)
    - Vessel elasticity (5 minutes)
    - Mental health/status or decline (takes money and time to access)
    - Dental health

Other Issues

- Continue to prioritize traits for which P3G will have *a priori* population data
- Consider benefits of having longitudinal data (re-contacting participants), such as the utility of studying long term effects of environment or medication on a population (note: in current state, CARTaGENE provides a snap shot of a genetically diverse population, without ability to follow up)
- Possibility of having sub-groups to get more details about a particular phenotype, if useful and feasible
- Possibility of obtaining medical information from initial physician and hospital
- Consideration of interactions between genetics and the environment will be important in order to analyze if predictive indicators of illness can be found (data should include social class, occupation, environmental exposure, for example to air pollution)
  - Consideration of the epidemiologist's perspective could be extremely useful
- Harmonization of sample analysis (QC, db, method) and need for flexibility within 'Federated Data Set'
  - Cores of P3G would be able to communicate with each other as a mechanism for control and cross talk, allowing comparisons between groups and ensuring quality control. It was suggested that the Core centres could have plates of blank forms that would be used to report back to P3G, therefore ensuring cross-talk
- Family component: cost-efficiency issues, encouragement of sib-pairs for participation, expansion of part of the samples to families
  - Addressed benefits of studying entire families, or asking individuals about incidence of disease in their family (note: Estonian study found that information given about family backgrounds is not always reliable)
  - Sample size may result in participation of siblings or spouse pairs, which could be an asset
  - Concerns raised about over-sampling in urban areas, and consideration of over-sampling rural areas, or keeping the over-sampled portion of the population in a separate pool to un-bias initial sample
  - Although follow-up is associated with a cost, it would make it possible to include additional parameters therefore harmonize various structures
- Tests must be feasible for a large scale study and quality control
- Issue of cost efficiency- evaluating methods and ensuring that funding covers cost of procedures

Biochemical measures (WG 5)

- **Method:** The group compiled a list of proposed biochemical measures and samples, specifically for the 'Federated Data Set' (from the UK Biobank, plus others that were relevant). The stability of analytes was checked in frozen samples to determine when measures should be taken on fresh samples. This was evaluated using published recommendations.

- **Biological samples:** The group proposed that only blood (fasting or not) and urine (spot or morning if possible) be collected and analyzed (fresh or frozen) for all participants. It was noted that it would be wise to collect both plasma and serum. Hair sampling could be considered in a subset of participants when mercury analyses are planned. Other samples mentioned in discussions were swabs and saliva (useful for periodontal disease, steroid hormone testing and assessing cigarette smoke exposure). The possibility of collecting lymphocytes for establishing immortalized cell lines was also discussed, although consent is needed to store blood samples and generate cell lines. Epigenetic markers (DNA methylation, redox status, urine nicotine, glyHb, carotene) were considered since they would provide quantitative data that is more useful than a questionnaire.

- **Biochemical parameters on fresh vs. frozen samples:** The main parameters to be measured on fresh samples are: CBC, Pt count, Vitamin C, HbA1c, RBC folate. This list was conditional upon approval by other members of the panel. Most of the other parameters listed in the UK Biobank can be measured on frozen samples. Sampling procedures, the anticoagulant, immediate sample handling and the addition of stabilizers may vary from analyte to analyte.

- **DNA damage:** The world-wide interest in the application of the micronucleus (MN) method to assess environmental effects on chromosome damage in blood and epithelial tissue was discussed. Combining available MN data from a variety of human populations would provide a powerful tool for evaluation of MN frequencies for public health and epidemiological studies. Performing this test would require isolating mononuclear cells and processing immediately. Scoring methods and results are described in *Mutat. Res*. 2003, 534(1-2):45-64.

- **Proteome:** Would likely want to separate three blood fractions which is technically demanding (must be done manually) and costly. Most proteins are stable when samples are frozen on liquid nitrogen, so they can be stored and measured in the future.

- **Transcriptome:** Difficult to purify mRNA in a standardized manner without affecting mRNA levels of individual genes. The best methods and materials to purify and preserve mRNA levels in blood cells of fractions must be discussed with laboratories having tested various approaches.

- **Freezing protocol and sample storage:** Issue must be further discussed as there are several available systems for managing and keeping frozen materials in a highly controlled and traceable way.

- **Common SOPs need to be written:** The mechanism of providing common SOPs needs to be defined. One proposed method would be to mandate a small working group to develop them starting from existing SOPs from each project.

- **Other issues**
  - Centralization of sample handling and analyses in a limited number of laboratories for each project
  - Standardization of methods, quality control and quality assurance of analyses
  - Determine analyses to be performed for all participants and those to be done on subsets of participants
  - Mechanisms by which a given P3G investigator will have access to samples to perform analyses in the context of limited quantities of materials.

Working Group 3: Storage, logistics and security

Leader: Professor Thomas J. Hudson

DNA storage:

- **Estonian Genome Project**
  - Centralized processing centre
  - 50 ml venous blood
  - DNA is extracted and salted out
  - Quality is checked by spectrophotometry, electrophoresis and one PCR reaction
  - Stored in liquid nitrogen

- **GenomEUtwin**
  - Centralized DNA extraction (Helsinki and Uppsala share a harmonized protocol)
  - DNA is extracted utilizing a Gentra Autopure workstation
  - Samples stored at -80ºC? liquid nitrogen?
  - Goal is to install robotized freezer system

- **UK Biobank**
  - Centralized processing centre
  - DNA measured by Pico Green automated system

- **CARTaGENE**
  - Genvault Robotic platform stores DNA dry on paper, 384-well format

- **Conclusions**
  - The group agreed that there was no need for harmonizing practices of DNA extraction, storage and management, as the goal of P3G was sharing of data, not samples
  - Paper method of storing DNA was discussed as better idea than storing DNA frozen, although this would not be imposed on all of the groups
  - Standards of DNA quantity and quality should be determined and met by each group
  - Quantity should be measured by spectrophotometry or the Pico Green system
  - In terms of DNA quality, standards should be set regarding DNA dilution protocols and PCR assays
  - Group discussed advantage of each project having a centralized processing centre

Non-DNA Storage:

- **Estonian Genome Project**
  - Plasma and buffy coats stored in liquid nitrogen

- **GenomEUtwin**
  - Blood samples have been collected by specific twin registries

- **UK Biobank**
  - Plasma and sera collected, frozen/dried and stored on paper
  - Buffy coats and extra plasma and sera stored in liquid nitrogen

- **CARTaGENE**

- Plasma and serum to be stored
- Monocytes will be frozen for potential transformation and immortalization

- **Conclusions**
  - Plasma/sera storage needs to be further discussed
  - Standards needed for the biological markers that will be measured

Logistics and Security

- **Estonian Genome Project**
  - Physician data encrypted (transportation code)
  - In processing centre, lab code is given
  - Final code kept by coding centre (links all information)
  - Coding centre for secure processing and storage of information, will connect all information on samples
  - Coding centre will not be connected to external network

- **GenomEUtwin**
  - Integration of six national twin cohorts
  - Sample information available to statistical core and individual twin cohort centre, but not genotyping core
  - Phenotype database will allow web-based data mining after data protection issues are addressed
  - Combining existing experiences with additional measures to develop security

- **UK Biobank**
  - Centralized processing and management of data
  - Group looking into various software
  - Oversight body for security

- **CARTaGENE**
  - Central recruitment office will handle initial contact, coding, consent, transfer to phenotyping centre, double-coding at the source of the recruitment process (RAMQ)
  - Phenotyping centres will be where questionnaires are completed on health, genealogy, demography, environment
  - Group will utilize *Entrust* genetic banking platform to coordinate data management
  - Data will also be stored by third-party computer company

Conclusions

- Group reiterated that although IT practices differed among projects, sharing of information and experiences would be helpful to all
- Need to create a readable, common language to allow various biobanks to communicate and share data
- Need to set standards for variables with regards to definitions and data formats (common nomenclature for disease codes)
- Communication standards should be developed (group discussed utilizing XML system to ensure standards of communication among the different projects)
- **Importance of adopting a common standard with regards to codes- for P3G, a single international organization should be responsible for handing out unique codes to samples, each beginning with a country code (ISO-standard), which will allow codes to be standardized and permit easy identification of their origin
- Harmonization of standards for secure data transfer
- Need to set up working groups, workshops and fund targeted studies

Other Issues

- Questionnaires and data entry- group discussed web questionnaires and need for new media such as digital paper and pen; language of questionnaires; method of administration (self or by GP)
- Coding for missing data
- Define data sharing, what and how?
- Ensure return of dataset to original biobank and archive results
- Rules and cost of access
- Will data be cross-sectional (CARTaGENE) or longitudinal (other projects)
- Cost due to differences in software
- Role of ICI- funding of research projects and central IT core

Working Group 4: Governance and ethical clearance

Leader: Professor Bartha Maria Knoppers

Governance/ethical clearance

Areas of Similarity

- Transparency
- Voluntary nature of participation
- Controlled access by third parties
- Openness towards sharing data with scientific community

Areas of Difference

- Nature and function of oversight bodies
- Nature of consent
- Level of identity and privacy protection
- Commercialization of data
- Right to withdraw
- Return of results
- Access by third parties

Possible areas of harmonization

- International oversight body to ensure standards for data-sharing, role and function needs to be discussed
- Core elements of consent, need to access rules of consent at the national level
- Level of privacy protection of the samples and data, need to assess impact of standardization on the individual projects
- Scientific integrity
- Conflict of interests

Commercialization

Areas of similarity

- Maximum public benefit
- Importance of academic freedom

Areas of difference

- Ownership of banks and samples
- Licensing
- Sharing of IP and copyright between industry, project, non-profit organizations and researchers
- Transfer of samples

Possible areas of harmonization

- Maximum public benefit world-wide
- Academic freedom and exchange of experts and young researchers
- Safeguard and ensure free flow of data while respecting national regulations and commercialization agreements
- Harmonizing standards of ownership and IP with respect to national regulations

Minimal Thresholds of Ethical Acceptance across all Projects

Recruitment

- No undue influence, coercion, financial inducement or overcompensation, misinformation, over-solicitation or immediate consent or invasion of privacy
- Recruitment must only be performed by persons with appropriate training
- Group agreed that differing methods of recruitment should not be an impediment

Consent

- How broad will consent be?
- Issue of consent vs. anonymization
- Ethically unacceptable: verbal or overly broad blanket consent, prohibition against international collaboration or overly exclusive research uses

Feedback to participants

- No unconsented feedback
- No unreasonable prohibition on dissemination

Right to withdraw

- No impractical right to withdraw
- Should be explained in the consent form

Destruction

- On-going stewardship
- No unnecessary abandonment of destruction of data without ethical clearance

Confidentiality/security

- No unauthorized or unnecessary third party access to identifiable data should be a principle
- Issue of anonymization- may prevent many benefits (CARTaGENE recognized need to address issue)
- No security measures should be incompatible with international collaboration

Transfer of data

- Need to establish contractual basis for access
- Projects should be ethically approved before data is released
- Data should be traced and sent back with results obtained (no secrecy of results)
- Exclusive access would be detrimental to P3G
- Issue of commercial use

Governance

- Governance should be addressed, and cannot be limited to ethical clearance
- No for-profit governance at P3G level
- No participation without disclosure of apparent or real conflicts of interest

Benefits sharing
- Members must be committed to the P3G goal of optimizing public benefit
- IP royalties need to be discussed
- Co-authorship should not be unreasonably restricted

Suggestions for an ideal dataset
- Representative sample
- On-going recruitment strategy
- Frequent exposure measures
- Virtual real-time outcomes
- Multi-centre recruitment
- Systematic variation in methodology
- Tangible outcomes
- Used for health service research, public health and health policy evaluation

Working Group 5: Public engagement

Leader: Alan Doyle

Background
- Lessons learned from public engagement prior/during project development
- Implementation of change/responses to public opinion

- **Estonian Genome Project**
  - For pilot project, provided information instead of advertising
  - Communication plan included articles in local press
  - Provided training and information material to data collectors
  - So far, project has been received positively by media, government, law and project

- **CARTaGENE**
  - Formed working and focus groups for public consultation
  - Web-site with information and opinions expressed in media
  - Meeting with government institutions
  - So far, population is in favor of the project, issues of confidentiality and security are sensitive
  - Next steps, focus groups, teleconferences, phone questionnaires

- **UK Biobank**
  - Consultation has taken place on the use of human tissue, and issues of collection and storage
  - So far, broad acceptance of project in principle
  - Public raised issues of recruitment, confidentiality, access to and uses of data, control and governance, value for money
  - Found misinformation and misunderstanding of genetic research
  - Consent and public ownership are crucial issues
  - Good communication with public is essential
  - Barriers to participation exist and solutions include web-site, focus group with public representatives
  - Integrated approach to information packages, presentations, briefing, communications strategy

- **Conclusions and key issues**
  - Issues similar across projects
  - Projects want to inform, not advertise
  - Positive public response dependent on quantity and understanding of information provided
  - Need to develop a long term strategic approach

Public engagement strategies

- Continuous monitoring and feedback of PR strategy/communications
- Models of current practice

- **Estonian Genome Project**
  - Need to provide information on genetics in general
  - Take into consideration level of comprehension of terminology and language
  - Planning integrated strategy that involves leaflets for clinics and schools
  - Recognized media and internet as key methods of communication
  - Evaluate public opinion over time with polls
  - Group wants to provide participants with research results, but do not have precise plan

- **CARTaGENE**
  - Aim to having public debate, and for information to circulate both ways between researchers and public
  - Integrated approach for communication including: web-site with forum, leaflets for clinics, drugstores and schools, media coverage
  - Group raised need to develop strategy for diffusion of scientific results by scientists

- **UK Biobank**
  - Standing focus group to stay in communication with public representatives
  - Utilizing special recruited team as well as Wellcome Trust and MRC communication teams
  - Integrated approach including different activities will be used

- **Conclusions and key issues**
  - Public engagement important since success of project is dependant on public participation
  - Need to address language used in communication strategies
  - Resources are needed to cover costs of communication strategies
  - Manage risk of popularization causing public concerns

Areas for convergence/standardization

- Benefits to harmonizing the messages
- Develop Q & A's to ensure all projects are on message
- Communication between projects about successful PR initiatives and communication experiences will lead to reduction in costs
- Promote that P3G is an international project, to give credibility and commit the public
- Opportunity for international approach to media coverage, including key identified journalists and an international spokesperson
- Advantages of developing international best practices
- Links between web-sites
- Harmonize message regarding pharmaceutical industry, that consortium is not promoted or driven by industry, but they do not want to close the door

Further Discussions

- Presence of an oversight body could promote public confidence and credibility
- Need for definition of 'the public', given that all parts of society should be involved in the process
- Feedback from participants would be of the greatest importance on issues of collection, uses, etc.
- Specificity of various projects and P3G consortium should be explained to participants
- Necessity to share tools for public consultation and engagement
- Use shared datasets for public purpose